

臉部特徵與眨眼檢測之專注力評估系統

應用於線上學習

Attention Evaluation System Based on Facial Features and Blink Detection Applying to Online Learning

劉育名¹ 徐豐明²

LIU, YU-MING¹ SHYU, FONG-MING²

¹ 臺中科技大學 多媒體設計系 研究生

¹ The Department of Multimedia Design of National Taichung University of
Science and Technology Graduate Student,

E-mail : yuming4142@gmail.com

² 臺中科技大學 多媒體設計系 副教授

² The Department of Multimedia Design of National Taichung University of
Science and Technology Associate Professor

E-mail : fms@nutc.edu.tw

摘要

2020 年因應 COVID-19 (新型冠狀病毒肺炎) 疫情席捲全球，許多出國留學的學生受到此影響無法在校進行學習，此波疫情賦予了遠距教學更多的應用，而這樣的影響卻也推動學術機構發展與投入新的線上課程資源。

遠距教學固然方便，與此同時疫情也帶給學生學習就業上的新挑戰，此時本論文所研究的透過影像分析判斷多數使用者專注力的系統就能夠達到實質上的幫助，使用者們經由視訊鏡頭以獲得臉部圖像，透過臉部特徵辨識與眨眼檢測綜合分析之後，讓授課者得到課堂中學生的專注力分析；也能夠達成使用者們本身的自我檢測，經由這些數據來得到評量與自我檢測的目的，經過實測，本研究達成 85.2% 的人臉辨識率再經由眨眼檢測獲得接近九成的專注度檢測準確率，適合應用於線上學習的檢測評估。

關鍵字：課堂觀察法、線上學習、專注度評估、臉部特徵辨識、眨眼檢測

ABSTRACT

In 2020, in response to the COVID-19 (new coronavirus pneumonia) epidemic that has swept the world, many students studying abroad have been affected by this and cannot study in school. This wave of epidemic has given a new definition of distance teaching, and this impact has promoted academics. The institution develops and invests in new online course resources. Although remote teaching is convenient, the epidemic also brings new challenges to students' learning and employment. At this time, the system studied in the paper that judges the concentration of most users through image analysis can achieve substantial help. Facial images have been obtained through the video lens, and after comprehensive analysis of facial feature recognition and blink detection, the faculty members can obtain the concentration analysis of the students in the classroom; it can also achieve the self-detection of the users themselves, and obtain the evaluation and self-detection through these data the goal of. This study achieved 85.2% of face recognition rate and then obtained nearly 90% of concentration detection accuracy through blink detection. The accuracy rate is suitable for the detection and evaluation of online learning..

Keywords: classroom observation, online learning, attention assessment, facial features recognition, blink detection

壹、前言

本研究是使用人工智慧(Artificial Intelligence, AI)與臉部特徵辨識(Facial features Recognition)的技術進行探討，結合上述技術來與使用者進行專注力的評估分析，經由臉部特徵辨識與眨眼檢測產生數據來呈現使用者在學習過程中的專注力，此過程透過睡意檢測、注視檢測與各方面的面部表情評估來達成，本章節共分四小節：研究背景、研究動機與目的、研究範圍與限制、研究流程與架構。

一、研究背景

由於人工智慧蓬勃的發展，應用到的層面相當的廣泛，如：醫療、金融、生產、監測等等，人工智慧尤其在影像處理、語音辨識和物聯網等等方面有相當大的成就。

人臉識別系統已經開發了近 30 年(Mita et al., 2005)。目前，人臉識別(Viola & Jones., 2004)技術已被引入到全球 40 個國家地區，並在世界上有 100 多個系統。由於對人工智能的研究而受到關注，例如深度學習和機器視覺(He et al., 2016)(Hearst et al., 1998)。2014 年，公司包括 Google，百度 IDL，微軟和 Facebook，像 Ginger 和 Face ++ (Levi & Hassner., 2015)這樣的初創公司也取得了不錯的成績。

課堂觀察是一種廣泛使用的工具，通常用於反饋給老師，並提高教育質量(Wragg., 2013)。典型的課堂觀察由觀察員參觀教室並進行錄音組成(Evertson & Burry., 1989)。課堂觀察系統通常在識別和評估教師對上課時間的使用以及教室中的互動類型。教室觀察系統局限性是存在反饋偏差的可能性，由於成本限制，觀察者一年只能兩次訪問一個教室，由於僱用和培訓觀察員的高昂費用，當前的課堂觀察做法在發展中國家構成了更大的挑戰。使用自動辨識系統代替觀察者。這將大大降低成本，並可能提高課堂觀察的真實性。

二、研究動機與目的

2020 年因應 COVID-19 (新型冠狀病毒肺炎) 疫情席捲全球，許多出國留學的學生受到此影響無法在校進行學習，此波疫情賦予了遠距教學新的定義。全球超過九成的實體課程受到影響甚至取消；教職人員的旅行或學術移動被迫暫停以及學術機構必須實施遠距工作，而這樣的影響卻也推動學術機構發展與投入新的線上課程資源。

遠距教學固然方便，與此同時疫情也帶給學生學習與就業上的新挑戰，此時本論文所研究的透過影像分析判斷多數使用者專注力的系統就能夠達到實質上的幫助，使用者們經由視訊鏡頭已獲得臉部圖像，透過臉部特徵與眨眼檢測綜合分析之後，讓教職者得到課堂中學生的專注力分析；也能夠達成使用者門本身的自我檢測，經由這些數據來得到評量與自我檢測的目的。

三、 研究範圍與限制

為了使研究的臉部圖像達到顯著的辨識效果，高質量穩定並具有高分辨率的鏡頭是必備的，此外照明的強弱也會影響辨識的精確度。因而本研究的範圍與限制可大致分為三部分：數據收集類型、系統設定部分與系統訓練階段。

(一)數據收集類型

將各資料庫獲取的圖像進行定位辨識與分類，分為各項臉部特徵的判定，並透過減少資料誤差值與盡可能提升精準度。

(二)系統設定部分

專注力檢測系統透過睡意檢測、注視檢測與各方面的面部表情評估，經由演算法得出陰性與陽性的結果來呈現，由於每個人的臉部特徵與表情管理是相當不一的，因此若要用客觀的判定使用者的行為是相當不簡單與不公平的事，因此在系統設定部分，數據異常的範圍，將以模型訓練出來的數據為主。

(三)系統訓練階段

此系統的前置作業是透過視訊鏡頭將影像中的圖像進行使用者臉部辨識系統，如公開資料集的訓練資料不就齊全，將會以研究者自行創造訓練及，此系統是以第三者的角度去觀察使用者進行評量的工作，可以應用在課程的學習評量與遠距教學的專注力分析。

(四)研究流程與架構

本研究分為五個階段來進行敘述，如圖 1.所示，依序為緒論、相關文獻探討、使用者資料蒐集、研究設備選擇與整合、預期結果。

1. **前言**：主要內容為探討研究背景與動機、確認研究目的，簡明研究範圍以及此研究方法的限制與如何解決的方法，並大致說明本研究的流程與架構。
2. **相關文獻探討**：本研究包含課堂觀察法的介紹，線上學習的相關文獻探討，專注力評估研究，演算法卷積神經網路 (Convolutional Neural Network, CNN) 的介紹和電腦視覺庫 (Open Source Computer Vision, OPENCV) 研究文獻，探討相關的文獻內容，選擇最適合臉部特徵辨識的方式，並敘述如何將其應用在本研究當中。
3. **使用者資料蒐集**：本章節將介紹使用此系統的使用情境，並詳細的介紹進行臉部辨識所要具備的硬體設備與外在環境的必要條件，最終透過數據呈現專注度與否的答案。

4. **研究設備選擇與整合**：本研究必須具備高接電子計算機與高質量穩定並具有高分辨率的鏡頭，將訓練資料透過 OPENCV 與 CNN 演算法進行機器學習與分類而後測試並修正。
5. **預期結果**：將訓練好的專注度評估系統進行使用者測試，進行一次辨識多張人臉透過其臉部特徵與眨眼檢測來分析其專注程度的情境測試，且逐步增加辨識的人臉數量，並保持一定的準確率。



圖 1. 研究流程與架構

貳、文獻探討

本研究是以人工智慧與臉部辨識為主軸，以實現本研究的臉部特徵與眨眼檢測之專注力分析系統。

一、 課堂觀察法

課堂觀察法是一種最直接，也最能深入瞭解老師教學狀況的一種諮詢方式，但需要老師與諮詢委員密切的合作，因為彼此雙方都扮演了重要的角色。在觀察前老師必須要提供當日教學單元的教材及教學目標，同時也需事先告知授課的班級，說明課堂觀察目的，表示尊重；觀察的當日，老師可以先向同學介紹一下諮詢委員，讓整個班級的氣氛不因為陌生者的介入而感到生疏。整個觀察的時間以一堂課為主，所以幾乎記錄了老師教學的全貌，非常適合欲全盤瞭解教學情況的老師。

黃瑞琴(2001)指出，依研究者介入被觀察者活動的程度，可以將研究者的角色區分為完全的觀察者、觀察者的參與 (an observer as participant)、參與者的觀察 (a participant as observer) 及完全的參與者 (a complete participant) 等四類。

類別	說明
完全的觀察者	觀察者在研究過程中居於旁觀者的地位，只觀察而不參與活動。
觀察者的參與	研究者完全參與，但須向研究對象表明自己觀察者的身份。
參與者的觀察	觀察者表明自己的身份，且和研究對象在社會過程中互動。
完全的參與者	研究者完全融入研究對象的生活中，且研究對象對方完全不知道研究者的身份。

圖 2. 觀察法的種類(鍾啟暘、林仁傑，2016)

然而課堂研究中經常使用的課堂觀察法，存在許多觀察偏差的問題，但其解決方法的討論卻不夠全面。陳向明(2001)即指出，在使用觀察法時，任何觀察活動都離不開觀察者的思考，都必須經過觀察者推論的過濾，因此我們需要有意識的對自己的推論進行反省，儘量做到將自己的推論和所觀察到的事情分開。然而這麼做卻十分的困難，如果無法做到如此，就有可能造成研究結果的偏誤。談到觀察者本身的因素，首先最常見的就是觀察者偏見、遺漏和偏移的問題。在個案研究、民俗志研究以及田野研究等研究中，研究者必須參與或投入研究參與

者的生活中，觀察其言行舉止，才能深入瞭解其行為深層意義，因此觀察研究法是經常使用的方法。而在教育研究中，大多數都要被研究者以語言或文字來表達，才能夠得到研究所需要資料，唯獨課堂觀察法利用觀察就能夠蒐集研究參與者的資料。也就因為觀察法有這樣的特性，所以使用觀察法的觀察者本身，就成為研究成果品質判定的重要關鍵。

觀察者的人格特質、價值觀等，都會影響觀察結果的分析與解釋，(葉重新，2005)而這就會造成觀察法中觀察偏差的第一個問題，也就是「觀察者偏見」。觀察者過去的經驗等主觀意識常會導致對情境、現象或行為產生不同的知覺和解釋，如果觀察者的偏見一旦產生，則觀察的結果將難以客觀公正。這種情況尤其容易發生在完全的觀察者 (a complete observer) 身上。

二、 線上學習

線上學習 (英語: Online Learning) 是一種透過網際網路工具來學習或訓練的方式。

線上學習是使用電子系統或電腦，以方便學習過程(Allen et al., 2016)。在線學習是一種利用互聯網和多媒體 (音頻, 視頻) 作為學習之間的傳遞，線上學習過程中主要媒介為導師和學生。因此，在線上學習教育系統在教學中利用基於互聯網的資訊技術來提高學生的知識和學習技巧。

線上學習的優勢和劣勢 Bates (2005)解釋說，線上學習提高了學生和教師之間的學習互動水平；促進學習與時間和地點靈活性的互動；能用實施於全球；易於更新內容以及保存資料的功能。但是，從過往研究來看，參加遠程線上學習課程的學生的出席率一直存在問題，與傳統的學習環境相比，輟學率異常高 (Richards & Ridley., 1997), (Wetzel., 1994)。30%至50%的輟學率被認為是相當普遍的(Moore & Kearsley., 2012)。與面對面的教育課程相比，學生在遠程在線學習中可能會感到孤獨感(Shaw & Polovina., 1999)。

三、 專注度評估

專注力，是人們完成一件事情或者學習新事物的重要基礎；以大腦功能來說，專注力並不只分佈在某些特定區域，而是分佈在廣泛的神經元網絡中，大腦會透過層級間或相互的回應將訊息連結在一起，並以此方式共同分工合作，當個體面對環境中許多刺激時，能選擇某些重要的訊息進行更詳盡的檢視，而忽略其他訊息的能力，稱之為注意力(attention)，就像是在吵雜的工作環境中，你仍然可以專注的進行文書處理，而忽略掉同事們的談話聲、音樂聲、腳步聲等。

專注力會隨著年紀發展，約持續到十二歲；一般性注意力的時間長度是三歲時約十五分鐘、六歲時約三十分鐘、九歲時約四十五分鐘，而高階的專注力(如：專心玩拼圖、積木、寫功課等)時間長度則是三歲約六分鐘、六歲約八分鐘、九歲約十分鐘(楊舒菁，2020)。

本研究是要將專注度評估系統應用於線上學習當中，以往的評估方法，如

2013 年 N H.Liu 等人(Han et al,2013)和 2015 年 C. M. Chen 等人(Chen et al,2015) 透過測量 EEG 腦波的方式,得到腦波特徵,並透過 Support Vector Machine(SVM) 演算法來評估學生的專注度,判斷學生是否處於專注狀態;2013 年 Raca 及 Dillenbourg(Raca & Dillenbourg,2013), 透過攝影機拍攝學生與教師,評估學生的視線方向與肢體動作來分析學生專注度;2017 年 Janez 及 Andrej(Zaleteli Andrej,2017)將專注度分為 5 個等級,並使用 Kinect 傳感器捕捉學生骨架加以進行特徵分類,再透過機器學系進行專注度評估;2020 年 Nirmal Krishnnan(Krishnnan et al,2020)等人提出的專注度評估方式,將臉部特徵透過睡意檢測、注視檢測和咀嚼檢測來進行分類,並使用二分法的方式呈現出使用者的專注度。

四、機器學習

機器學習(Machine Learning)的應用非常的廣泛,例如;物體識別、語音識別、聲波探測、自然語言處理、語意分析、預測、語言翻譯、人臉辨識、物體辨識...等。

機器學習是透過演算法將收集到的資料進行分類或預測模型訓練,在未來中,當得到新的資料時,可以透過訓練出的模型進行預測。機器學習是一種弱人工智慧(narrow AI),它可以讓電腦尋找資料中得到複雜的函數(或樣本)來學習並創造演算法(或一組規則),並利用它來做預測。同時並藉由驗證資料比對計算分類結果,來判定模型是否適合用來預測或分類。

機器學習可以分成下面幾種類別：

- 監督學習：從給定的訓練資料集中學習出一個函式，當新的資料到來時，可以根據這個函式預測結果。監督學習的訓練集要求是包括輸入和輸出，也可以說是特徵和目標。訓練集中的目標是由人標註的。常見的監督學習演算法包括回歸分析和統計分類。
- 無監督學習：訓練集沒有人為標註的結果。常見的無監督學習演算法有生成對抗網路（GAN）。
- 半監督學習：介於監督學習與無監督學習之間
- 增強學習機器：為了達成目標，隨著環境的變動，而逐步調整其行為，並評估每一個行動之後所到的回饋是正向的或負向的。

(一)卷積神經網路

卷積神經網路(Convolutional Neural Network, CNN)是一種前饋神經網路，它的人工神經元可以回應一部分覆蓋範圍內的周圍單元，(24)對於大型圖像處理有出色表現。

卷積神經網路由一個或多個卷積層和頂端的全連通層（對應經典的神經網路）組成，同時也包括關聯權重和池化層（pooling layer）。這一結構使得卷積神經網路能夠利用輸入資料的二維結構。與其他深度學習結構相比，卷積

神經網路在圖像和語音辨識方面能夠給出更好的結果。這一模型也可以使用反向傳播演算法進行訓練。相比較其他深度、前饋神經網路，卷積神經網路需要考量的參數更少，使之成為一種頗具吸引力的深度學習結構 (<https://reurl.cc/OX15GD>)。

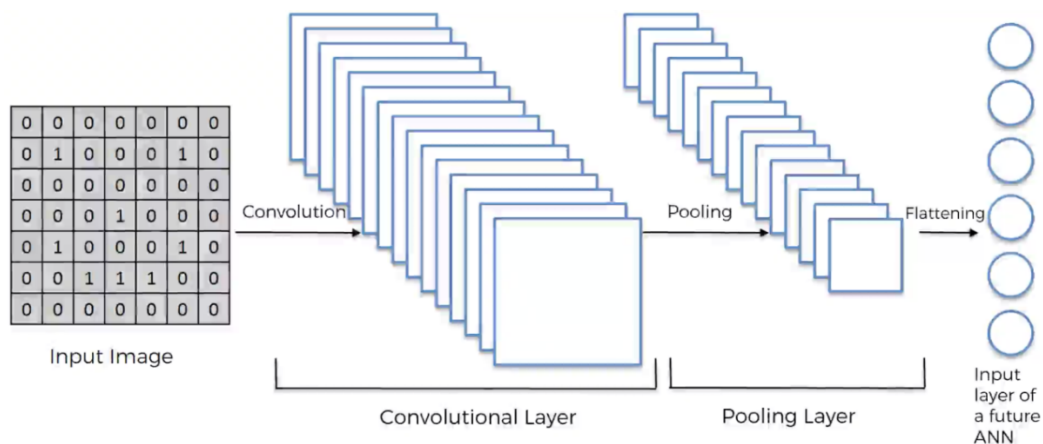


圖 3. 卷積神經網路 (Convolutional Neural Network, CNN)
(資料來源: <https://reurl.cc/003qqA>)

1. 池化層

池化層(Pooling Layer)在卷積神經網路 (Convolutional Neural Networks) 扮演的角色也很關鍵，它可以幫助我們縮小 Feature map 的大小 (也就是降維)，也可以用來強 CNN 萃取出來的特徵。池化 (Pooling) 是卷積神經網路中另一個重要的概念，它實際上是一種非線性形式的降採樣。有多種不同形式的非線性池化函式，而其中最大池化 (Max pooling) 是最為常見的。它是將輸入的圖像劃分為若干個矩形區域，對每個子區域輸出最大值。直覺上，這種機制能夠有效的原因在於，在發現一個特徵之後，它的精確位置遠不及它和其他特徵的相對位置的關係重要。池化層會不斷地減小數據的空間大小，因此參數的數量和計算量也會下降，這在一定程度上也控制了過擬合。通常來說，CNN 的捲積層之間都會周期性地插入池化層(黃小天、李亞洲，2017)。

常見的池化動作是 Max pooling，所以一個 4x4 的 feature map，經過 Max-pooling 之後就會形成 2x2 的矩陣(如圖 4.)。

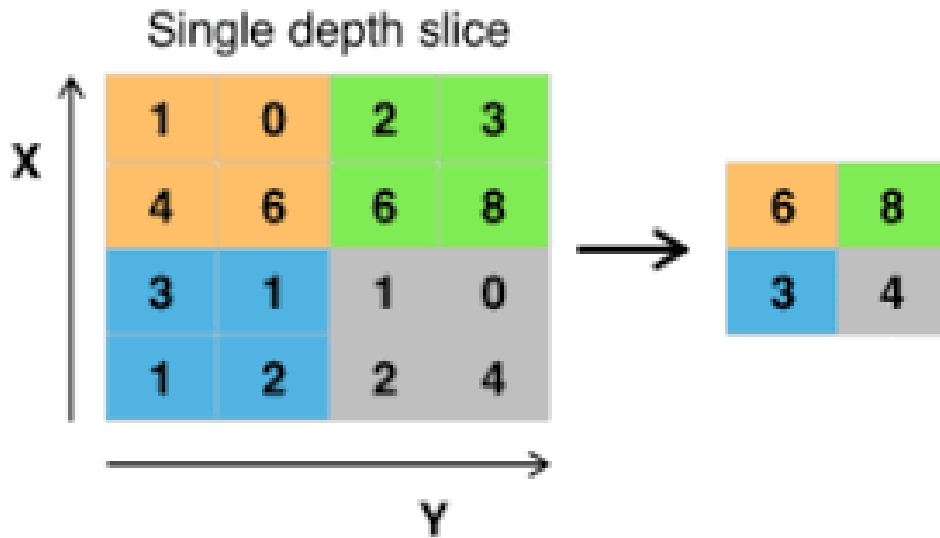


圖 4. 池化窗口為 2x2 的最大池化層(資料來源: 機器之心(黃小天、李亞洲, 2017))

2. 平坦層

平坦層 (Flatten Layer)在 CNN 前面幾層都是卷積層跟池化層交互轉換，後半段會使用多層感知器來穩定判斷結果。所以再接入多層感知器前，先必須將矩陣打平成一維的陣列作為輸入，然後再串到後面的隱藏層跟輸出層。

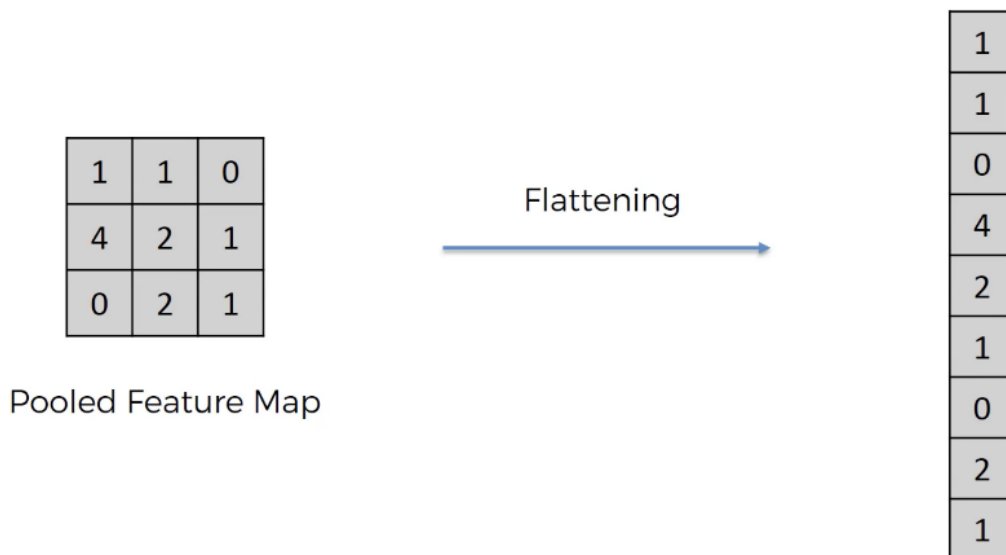


圖 5. 平坦層(資料來源: <https://reurl.cc/0O3Vmx>)

3. 整流線性單位函數 ReLU

線性整流函數 (Rectified Linear Unit, ReLU), 又稱修正線性單元, 是一種人工神經網絡中常用的激勵函數 (Activation Function)。

神經網絡模型中, 由於輸出和輸入依然脫離不了線性關係, 所以各隱藏層、包括輸出層都需要激勵函數, 激勵函數能夠幫助我們引入非線性因素, 使得神經網絡能夠更好地解決更加複雜的問題。

截至目前為止, 在深度學習領域 ReLU 激勵函數蔚為主流, 主要考量的因素有以下幾點:

- (1) 梯度消失問題 (Vanishing Gradient Problem):

ReLU 的分段線性性質能有效的克服梯度消失的問題。對使用反向傳播練的類神經網絡來說, 梯度的問題是最重要的。

- (2) 類神經網路的稀疏性 (奧卡姆剃刀原則):

Relu 會使部分神經元的輸出為 0, 可以讓神經網路變得稀疏, 緩解過度擬合的問題。

- (3) 生物事實: 全有全無律 (All or None Law):

在神經生理方面, 當刺激未達一定的強度時, 神經元不會興奮, 因此不會產生神經衝動。如果超過某個強度, 才會引起神經衝動。ReLU 比較好的捕捉了這個生物神經元的特徵。

- (4) 計算量節省: Relu 計算量小, 只需要判斷輸入是否大於 0, 不用指數運算。

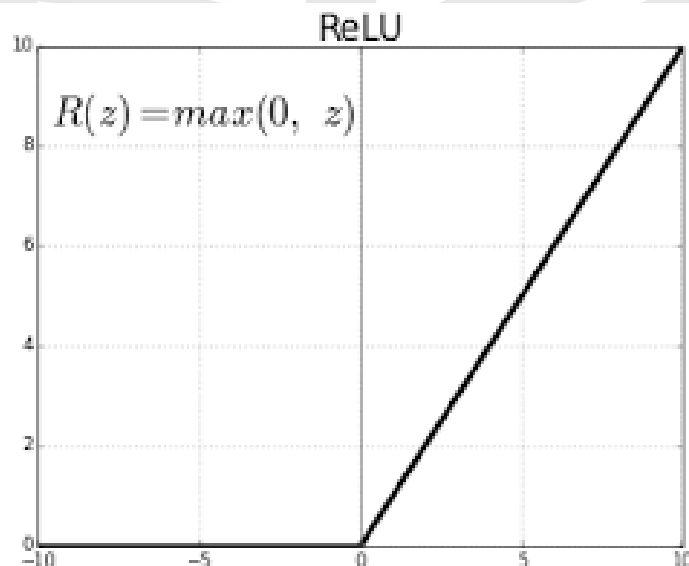


圖 6. 整流線性單位函數 ReLU (資料來源: <https://reurl.cc/main/tw>)

4.分類器

Softmax 分類器是深度學習領域相當常見的一個分類器，Softmax 分類器可以當成一個多分類的邏輯回歸。

Softmax 函數，又稱歸一化指數函數。Softmax 函數通常會放在類神經網路的最後一層，將最後一層所有節點的輸出都通過指數函數(exponential function)，並將結果相加作為分母，個別的輸出作為分子。

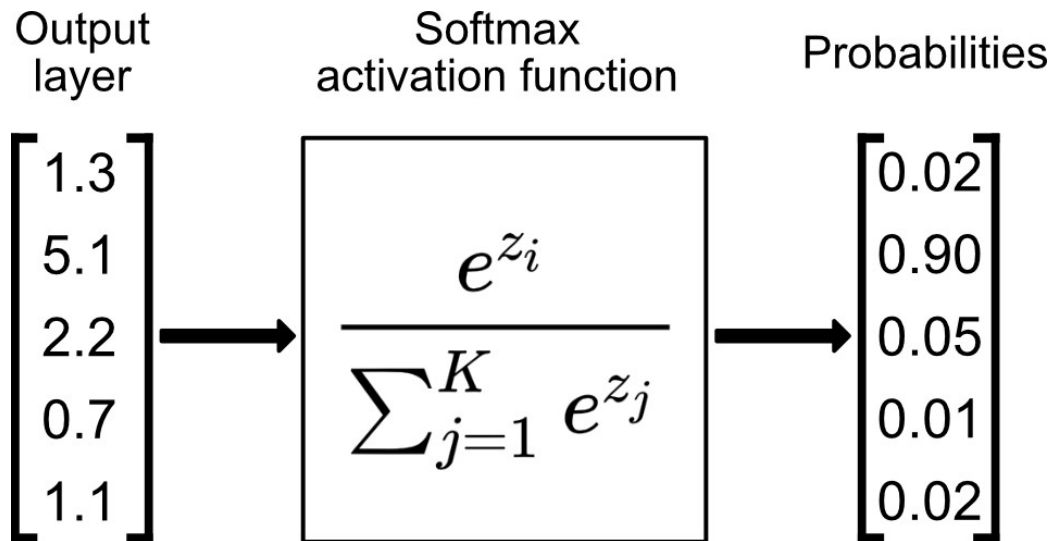


圖 7. Softmax 歸一化 (資料來源: <https://reurl.cc/00AgQA>)

(二)人臉偵測與識別

人臉辨識 (face recognition) 的應用廣泛，甚至有人認為人臉辨識的發展，將可能成為改變人類生活模式的一大關鍵，因為近年不只是科技大廠關注這一塊的發展，現在金融、零售、醫療、安防等領域的企業與政府單位也正做好準備，並已有導入的案例，利用人臉辨識技術來確認身份、簡化人力、蒐集數據與提供服務，還有能用於照護與犯罪偵防的場景之中

人臉檢測 (face detection) 是一種在任意數字圖像中找到人臉的位置和大小的計算機技術。它可以檢測出臉部特徵，並忽略諸如建築物、樹木和身體等其他任何東西。有時候，人臉檢測也負責找到面部的細微特徵，如眼睛、鼻子、嘴巴等的精細位置。

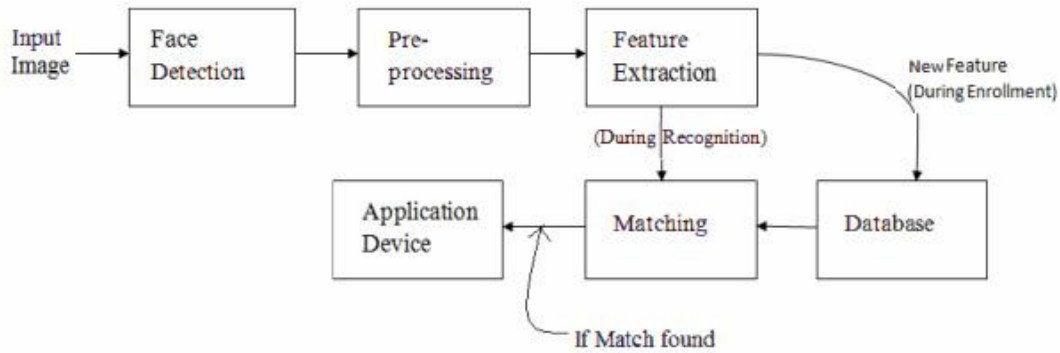


圖 8. 人臉識別系統架構圖(資料來源: (Krishna et al,2014))

(三) 臉部特徵偵測

面部辨識取得了很大的進步過去幾年的認可。越來越多的系統即使在不受限制的情況下也能獲得出色的性能在以下支持下，環境不再是問題硬件和算法。著名的標籤臉野生 (LFW) (Huang et al,2007)數據集包含具有不同面部圖像的面部圖像姿勢，面部表情和照度的變化通常用作比較基準面部識別系統的研究。擁有 97.35% 識別準確性，Facebook 的 DeepFace 是第一個系統達到接近人類水平的性能 (97.53%)。

五、 OPENCV 介紹

電腦視覺資料庫(OpenCV)全名是 Open Source Computer Vision Library，是當今最知名、也最被廣泛採用的影像處理函式庫，它是由 Intel 發起並參與開發，以 BSD 授權條款發行，可在商業和研究領域中免費使用，目前是非營利的基金組織 OpenCV.org 進行維護。

OpenCV 用 C 語言和 C++ 語言編寫，可以在 Windows，Linux，Mac OS X 等系統運行。同時同時積極開發 Python，Java，Matlab 以及其他一些語言的接口，將庫引入安卓和 iOS 中為移動設備開發應用

OpenCV 庫包含從計算機視覺各個領域衍生出來的 500 個多個函數，包括工業產品質量檢驗，醫學圖像處理，安保領域，交互操作，相機校正，雙目視覺以及機器人學。因為計算機視覺和機器學習經常在一起使用，所以 OpenCV 也包含一個完備的，具有通用性的機器學習庫 (ML 模塊)。這個子庫聚焦於統計模式識別以及聚類。ML 模塊對 OpenCV 的核心任務 (計算機視覺) 相當有用，但是這個庫也足夠通用，可以用於任意機器學習問題。

(一) Haar 分類器

Haar 分類器用到了 Boosting 演算法中的 AdaBoost 演算法，只是把 AdaBoost 演算法訓練出的強分類器進行了級聯，並且在底層的特徵提取中採用了高效率的矩形特徵和積分圖方法。

哈爾特徵 (英語: Haar-like features) 是用於物體辨識的一種數位影像特徵。它們因為與哈爾小波轉換極為相似而得名，是第一種即時的人臉檢測運算。

過去，直接使用圖像的強度（就是圖像每一個像素點的 RGB 值）使得特徵的計算強度很大。Papageorgiou 等人提出可以使用基於哈爾小波的特徵而不是圖像強度(Papageorgiou& Poggio,1998)。Paul Viola and Michael Jones(Viola & Jones,2001)進而提出了哈爾特徵。哈爾特徵使用檢測窗口中指定位置的相鄰矩形，計算每一個矩形的像素和並取其差值。然後用這些差值來對圖像的子區域進行分類。

矩形 haar 特徵

一個矩形哈爾特徵可以定義為矩形中幾個區域的像素和的差值，可以具有任意的位置和尺寸。這種特質也被稱為 2 矩形特徵 (2-rectangle feature)。維奧拉和瓊斯也定義了 3 矩形特徵和 4 矩形特徵。這個值表明了圖像的特定區域的某些特性。每一個特徵可以描述圖像上特定特性的存在或不存在，比如邊緣或者紋理的變化。例如，一個 2-矩形特徵可以表明一個邊界位於一個亮區域和一個暗區域之間。

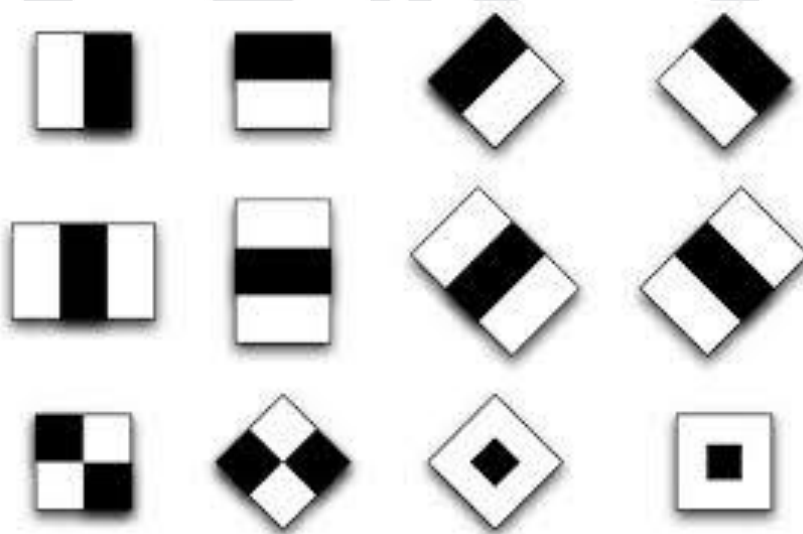


圖 9. 矩形 haar 特徵(資料來源: <https://reurl.cc/KjNpQm>)

(二)HSV 顏色模型

HSV 模型通常用於計算機圖形應用中。在用戶必須選擇一個顏色應用於特定圖形元素各種應用環境中，經常使用 HSV 色輪。在其中，色相表示為圓環；可以使用一個獨立的三角形來表示飽和度和明度。典型的，這個三角形的垂直軸指示飽和度，而水平軸表示明度。在這種方式下，選擇顏色可以首先在圓環中選擇色相，在從三角形中選擇想要的飽和度和明度。

HSV 模型的另一種可視方法是圓錐體。在這種表示中，色相被表示為繞圓錐中心軸的角度，飽和度被表示為從圓錐的橫截面的圓心到這個點的距離，明度被表示為從圓錐的橫截面的圓心到頂點的距離。某些表示使用了六角錐體。這種方法更適合在一個單一物體中展示這個 HSV 色彩空間；但是由於它的三維本質，它不適合在二維計算機界面中選擇顏色。

HSV 色彩空間還可以表示為類似於上述圓錐體的圓柱體，色相沿著圓柱體的外圓周變化，飽和度沿著從橫截面的圓心的距離變化，明度沿著橫截面到底面和頂面的距離而變化。這種表示可能被認為是 HSV 色彩空間的更精確的數學模型；但是在實際中可區分出的飽和度和色相的級別數目隨著明度接近黑色而減少。此外計算機典型的用有限精度範圍來存儲 RGB 值；這約束了精度，再加上人類顏色感知的限制，使圓錐體表示在多數情況下更實用。

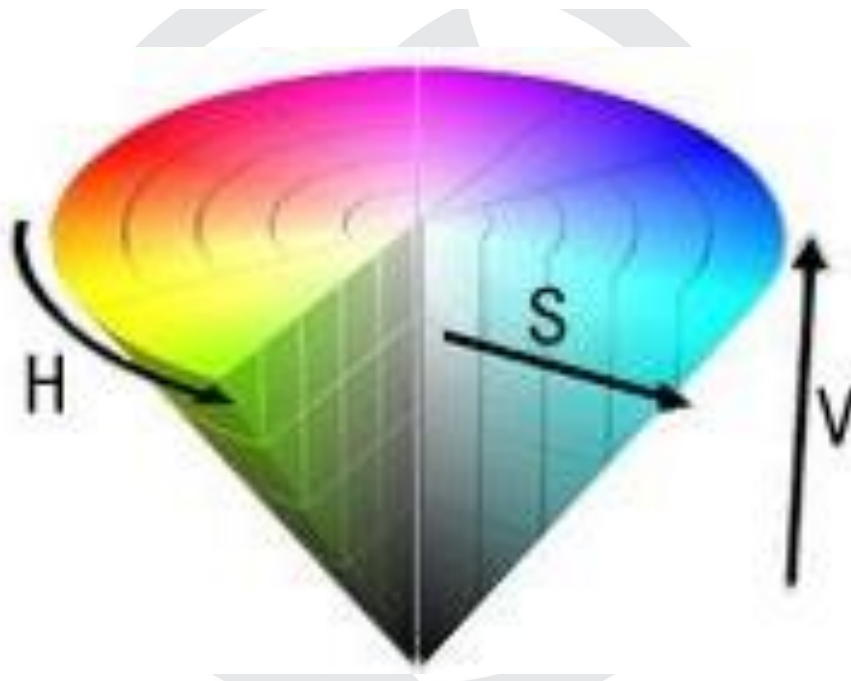


圖 10. HSV 顏色空間模型(資料來源: <https://reurl.cc/j5A0eq>)

(三)RGB 顏色模型

顏色通常用三個獨立的屬性來描述，三個獨立變量綜合作用，自然就構成一個空間坐標，這就是顏色空間。但被描述的顏色對象本身是客觀的，不同顏色空間只是從不同的角度去宣稱同一個對象。

計算機色彩顯示器和彩色電視機顯示色彩的原理相同，都是採用 R，G，B 相加混色的原理，通過發射出三種不同強度的電子束，使屏幕覆蓋的紅，綠，藍磷光材料發光而產生色彩。這種色彩的表示方法稱為 RGB 色彩空間表示。在 RGB 顏色空間（如圖 1.1）中，任意色光 F 都可以用 R，G，B 三色不同分量的相加混合而成： $F = r(R) + r(G) + r(B)$ 。RGB 色彩空間還可以用一個三維

的立方體來描述。當三基色分量都為 0（最弱）時混合為黑色光；當三基色都為 k （最大，值由存儲空間決定）時混合為白色光。RGB 色彩空間採用物理三基色表示，因此物理意義很清楚，適合彩色顯像管工作。而這一機構並不適應人的視覺特點。由此，產生了其他不同的色彩空間表示法。

(四) 圖像二值化

二值化（英語：Binarization）是圖像分割的一種最簡單的方法。二值化可以把灰度圖像轉換成二值圖像。把大於某個臨界灰度值的像素灰度設為灰度極大值，把小於這個值的像素灰度設為灰度極小值，從而實現二值化。

二值化影像即是影像中只有「黑」與「白」的表現，將一個彩色影像轉為灰階影像後，針對影像中的每個像素點，設定一個灰階值標準，也稱作閾值(threshold)，灰階值大於標準的像素點轉為白點，灰階值小於標準的像素點轉為黑點，經過轉換就可以得到一張二值化影像。(如果是 RGB 彩圖，則需要先將像素點的 RGB 轉成灰度值)設置為 0 或 255，也就是將整個圖像呈現出明顯的黑白效果的過程。



圖 11. 圖像二值化(資料來源: <https://reurl.cc/EzAdkv>)

參、研究實施與設計

依據文獻探討設計本研究之系統與實驗內容規劃，以下由第一小節開始說明研究步驟與研究方法和系統架構與流程，以及介紹體驗具體事項，在依據此流程進行設計與開發。

一、研究流程

本研究開始時會收集相關的文獻與資料，去得知目前專家學者們對於研究臉部特徵與眨眼辨識所使用的方法與碰到的問題，並經由比較其優劣勢來決定研究的方式與確定系統的架構，系統架構執行的部分會先透過蒐集各大人臉識別的公開資料庫經由 Opencv 去進行臉部偵測，並提取其中的臉部特徵包含頭部，眼部，嘴部，並採用機器學習 CNN 去進行臉部特徵分類，並讓機器學習並訓練出能夠識別本研究所需的模型，並依分類來達成專注度相關的評估，而後讓系統一次性識別多張臉部來計算出多數人專注力的分析。

二、研究方法

研究方法我們將其分為四大步驟：

- 1.系統分析與前置作業準備階段：文獻蒐集、圖像提取、圖像預處理
- 2.系統研發階段：人臉檢測、臉部特徵提取、眨眼檢測
- 3.功能開發與測試階段：臉部特徵的精確率分析、使用者專注度精確率分析
- 4.系統測試及修正階段：多人專注度評估、提高準確率

(一)析與蒐集相關的文獻與資料

本研究會蒐集研究所需的相關文獻，從國外網站和著作與校內外圖書館館藏中，尋找有關於本研究的相關資料，整理出對此有益的資訊，並找出其相關領域所不足的部分，以著重研究。關於系統分析的部分，將透過研究過的相關文獻，尋找出最有益於此系統的建構方式。

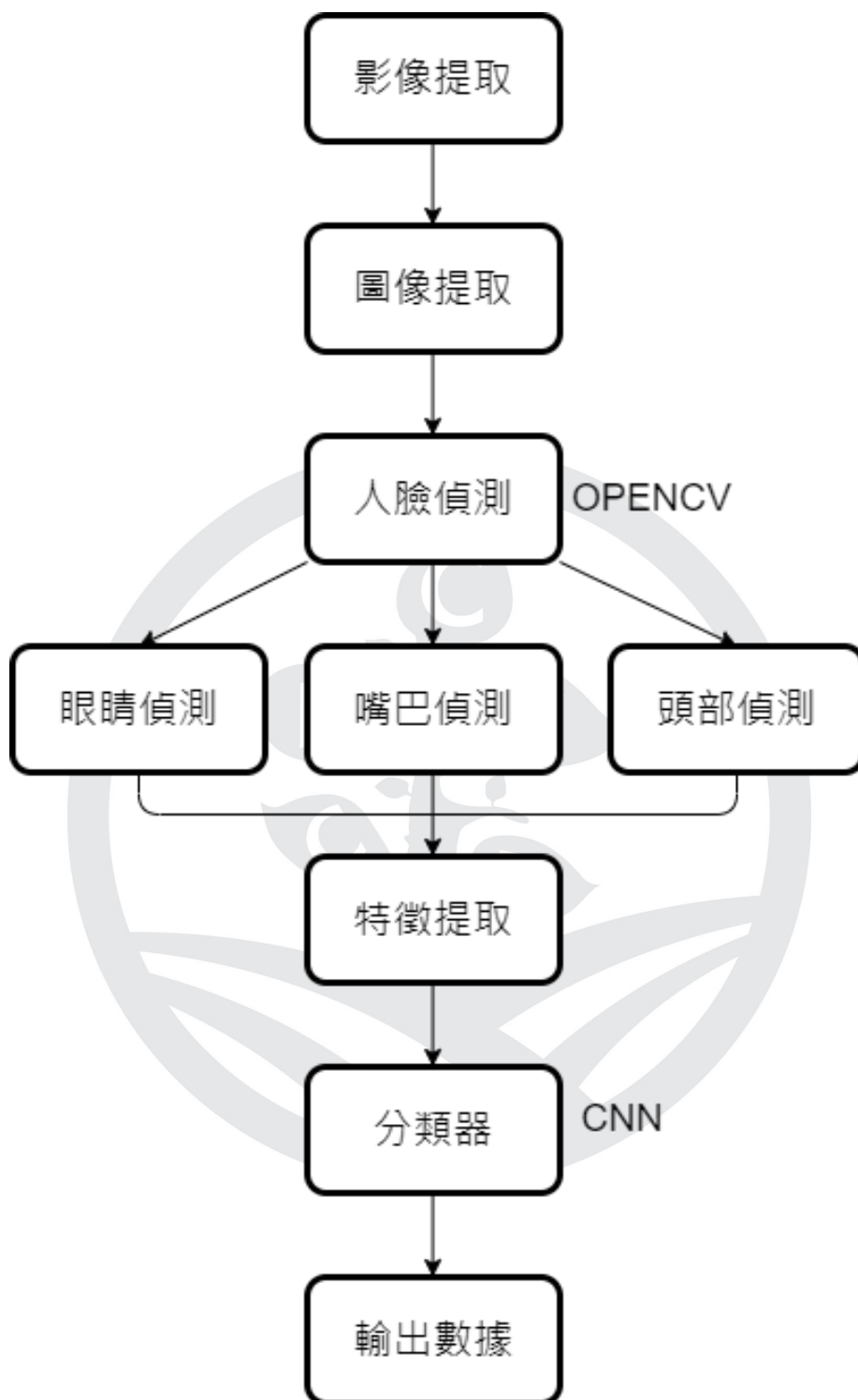


圖 12. 訓練模型階段

(二)系統研發設計階段

本系統由四個主要架構所組成—機器學習、人臉辨識、臉部特徵提取，臉部特徵分類，結合四種技術針對目標族群為課堂中學習的學生進行測試，透過影片中所獲取的圖像，我們可以透過學生們的眼睛檢測、嘴唇檢測和頭部仰角檢測來判斷其專注力的程度。

前處理的部分透過提取影片中的圖像，再將提取好的圖像進行人臉辨識，臉部偵測與特徵分類部分是採用 OPENCV、膚色分割法、Haar 特徵分類器...等方式，將各項特徵分類後透過 CNN 進行訓練，透過消融實驗找出最佳函數，並訓練出效果最為顯著的模型，而後分析多人專注度測試並將其應用在遠距教學的課堂觀察中。

(三)功能開發與測試階段

功能開發部分會參照 Nirmal Krishnnan(2020)等人提出的專注度評估方式，將臉部特徵透過睡意檢測、注視檢測和咀嚼檢測來進行分類，並使用二分法的方式呈現出使用者的專注度。系統開發會透過提取訓練資料影片中的圖像，再將圖像透過 Opencv 將人臉定位，透過 CNN 演算法使機器學習並分類訓練，而後將模型進行精確度分析，再透過使用者實測並逐步增加一次測試的人數，並分析其模型準確率和學生們的專注力。

(四)系統測試及修正階段

對於多數人專注度評估系統的開發，臉部特徵辨識的精確度是首先最必要滿足的條件，而後逐步增加辨識人數，並保持一定的精準度以達成多人專注度的評估。

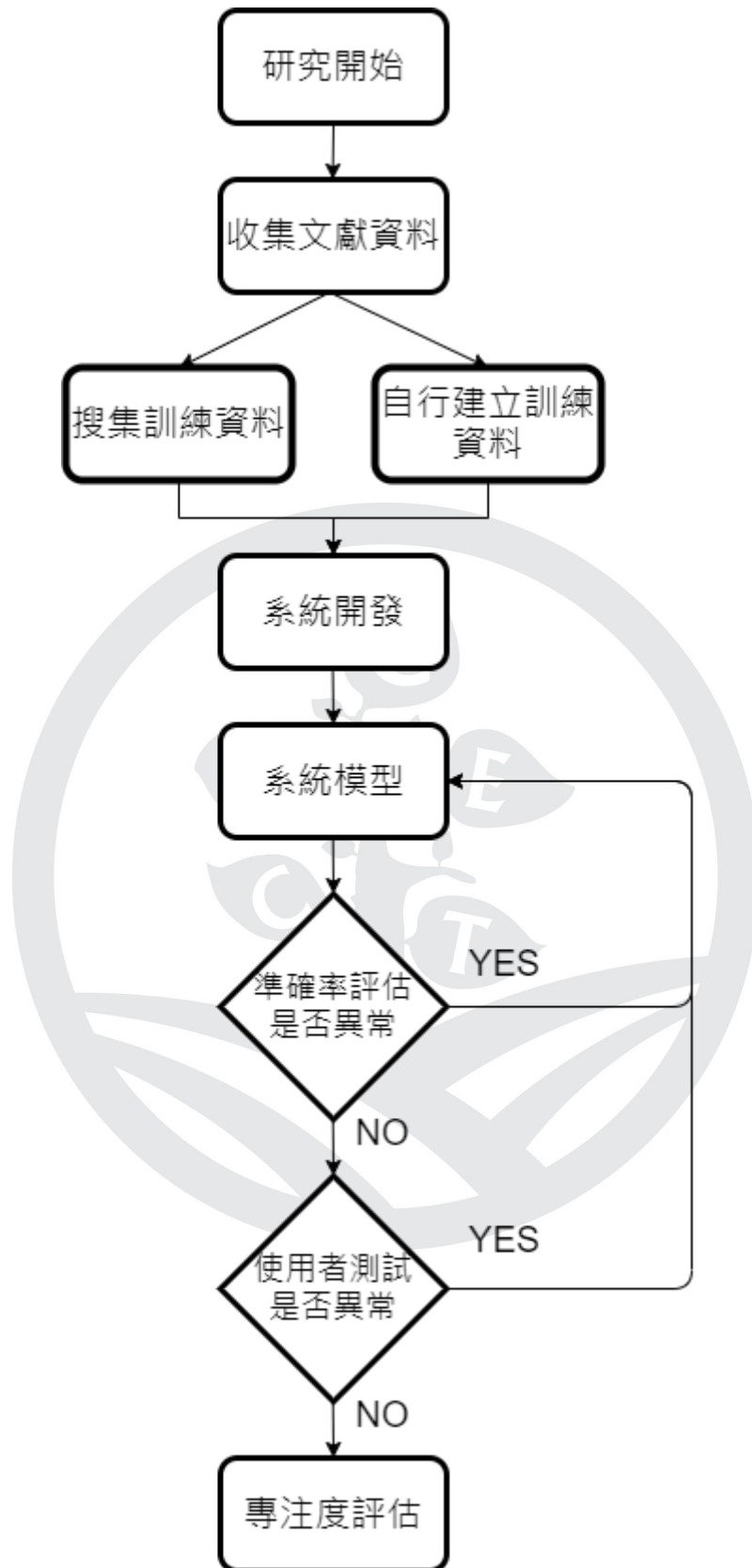


圖 13. 系統流程與架構

肆、 結果與討論

此專注力評估系統經過實測，達成 85.2% 的人臉辨識率再經由眨眼檢測獲得接近九成的專注度檢測準確率，適合應用於線上學習的檢測評估，檢測結果如下：

一、 效能評估

本研究透過自製資料集蒐集了十名自願參與本研究的同學，經過影像識別技術提取人臉圖片進行臉部偵測與辨識，將 250 張人臉圖片透過預處理分為 175 張訓練圖片與 75 張測試圖片進行臉部辨識，達成 85.2% 的辨識準確率。

人臉偵測的結果根據檢測的結果可以將其分成以下四種：

1. TP(True Positive): 正確預測成功的正樣本，
2. TN(True Negative): 正確預測成功的負樣本，
3. FP(False Positive): 錯誤預測成正樣本，實際上為負樣本
4. FN(False Negative): 錯誤預測成負樣本

		實際情形	
		人臉	非人臉
偵測情形	人臉	TP	FP
	非人臉	FN	TN

圖 14. 人臉偵測情形

了解到以上檢測可能的情形，我們就可以依據 TN, FP, FN, TP 計算各式比率，以衡量模型的效能，如表 1：

準確率 (Accuracy) = $(TP+TN)/(TP+FP+FN+TN)$

精確率 (Precision) = $TP/(TP+FP)$ ，即陽性的樣本中有幾個是預測正確的。

召回率 (Recall) = $TP/(TP+FN)$ ，即事實為真的樣本中有幾個是預測正確的。

F1 = $2/((1/Precision) + (1/Recall))$ ，即精確率與召回率的調和平均數。

表 1. 人臉辨識結果

人臉總數 250	TP	FN	FP	準確率	精確率	錯誤率
本研究結果	213	37	35	85.2%	85.9%	15%

二、 眨眼檢測

本研究採用自製資料集，將眨眼次數每分鐘超過 40 次、視線離開螢幕、人臉離開螢幕視為不專心。錄製了十名同學遠距學習的十分鐘短片，並將其切割成數個一分鐘的影片以便分類為專心與不專心的學習情況，將影片分為 150 部作為訓練資料與 50 部當作驗證資料，而後透過機器學習判斷使用者的學習專注度，結果如表 2、表 3 所式：

表 2. 眨眼檢測結果

眨眼次數	辨識正確	辨識錯誤	準確率
每分鐘介於 20-40 次	28	4	87.5%
每分鐘大於 40 次	15	3	83.3%

表 3. 專注度檢測結果

行為分類	辨識正確	辨識錯誤	準確率
專心	26	2	92.9%
不專心	19	3	86.4%

伍、 未來展望

本研究提出一個透過眼部去觀察使用者是否專注的方法，由於眼球偵測極易受到障礙物抑或是光照的影響而造成遮蔽而無法偵測，故本研究只考慮將視訊鏡頭擺在人臉正前方的狀況，未來的目標是研究改良人臉特徵的抽取方式，讓其不必受到角度與光照影響導致偵測率的下降。

在專注度辨識的方面，由於樣本數量稀少與是否具有代表性是會直接影響到此系統的可信度。然而由於資料集影片取得不易，故訓練樣本與測試樣不夠充足，因此在後續的研究當中，首要目標是提升資料集的數量，並花費更多時間蒐集樣本母體並且透過更多方面的檢測方式，並將其應用在生活中，達成遠距學習專注度檢測的目的。

參考文獻

一、中文部分

- 黃瑞琴 (2001), 質的教育研究方法。臺北：心理出版社，38-40。
- 鍾啟暘、林仁傑 (2016), 教育研究中課堂觀察法的觀察偏差問題探究-以體育課為例，28-35。
- 陳向明 (2001), 教師如何做質的研究。北京：教育科學出版社，70-78。
- 葉重新 (2005), 教育研究法。臺北：心理出版社，42-44。
- 楊舒菁(2020)。不可不知的專注。 <http://ttw3.mmh.org.tw/hospital/paper19.html> ,
上網日期:2021/1/9。
- 黃小天和李亞洲,【機器之心】人工智能與自然語言處理概述:AI 三大階段、NLP 關鍵應用領域，
上網日期:2021/1/7。

二、英文部分

- T. Mita, T. Kaneko and O. Hori. (2005), *Joint haar-like features for face detection*, Tenth IEEE International Conference on Computer Vision, 1619-1626.
- P. Viola and M. J. Jones. (2004). *Robust real-time face detection*, International journal of computer vision, vol. 57, no. 2, 137-154.
- K. He, X. Zhang, S. Ren and J. Sun. (2016). *Deep Residual Learning for Image Recognition*, "IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.
- M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf. (1998). *Support vector machines*, IEEE Intelligent Systems and their Applications ,vol: 13.
- G. Levi and T. Hassner. (2015). *Emotion recognition in the wild via convolutional neural networks and mapped binary patterns*, in Proceedings of the ACM on international conference on multimodal interaction, 503-510.
- T. Wragg. (2013). *An Introduction to Classroom Observation* (Classic Edition). Routledge.
- C. M. Evertson and J. A. Burry. (1989). *Capturing classroom context: The observation system as lens for assessment*, Journal of Personnel Evaluation in Education, vol. 2, no. 4, 297-320.
- M. Allen, Michael Allen. (2013). *guide to e-learning*. Hoboken: John Wiley.
- T. Bates, Technology. (2005). *e-learning and distance education*. London: Routledge.
- C. Richards and D. Ridley. (1997). *Factors affecting college students' persistence in on-line computer-managed instruction*, Coll. Stud. J., vol. 31, no. 4, 490-495.
- C. Wetze. (1994). *Instructional effectiveness of video media*. Lawrence Erlbaum Associates, Inc.

- M. Moore and G. Kearsley. (2012). *Distance education*. Belmont, CA: Wadsworth Cengage Learning.
- S. Shaw and S. Polovina. (1999). *Practical Experiences of, and Lessons Learnt from, Internet Technologies in Higher Education*, J. Educ. Technol. Soc., vol. 2, no. 3, 16–24..
- L. N. Han, C. Y. Chiang, and H. C. Chu. (2013). *Recognizing the Degree of Human Attention Using EEG Signals from Mobile Sensors*, Sensors, vol. 13, no. 8, 10273-10286.
- C. M. Chen, J. Y. Wang, and C. M. Yu. (2015). *Assessing the Attention Levels of Students by a Novel Attention Aware System based on Brainwave Signals*, 2015 IIAI 4th International Congress on Advanced Applied Informatics, 379-384.
- M. Raca and P. Dillenbourg. (2013). *System for Assessing Classroom Attention*, Proceedings of the Third International Conference on Learning Analytics and Knowledge, 265-269.
- J. Zaletelj and A. Košir. (2017). *Predicting students' attention in the classroom from Kinect facial and body features*, EURASIP Journal on Image and Video Processing, 1-12.
- Nirmal Krishnnan, Saeed Ahmed, Thanmay Ganta and Gurusamy Jeyakumar. (2020). *A Video Analytics Based Solution for Detecting the Attention Level of the Students in Class Rooms*, IEEE Conference on International Conference on Cloud Computing, Data Science & Engineering (Confluence).
- Convolutional Neural Networks (LeNet) - DeepLearning 0.1 documentation.
DeepLearning 0.1. LISA Lab. <https://reurl.cc/OX15GD> , 上網日期:2021/1/7.
- Convolutional Neural Network. (2014).
<https://blog.xuite.net/metafun/life/589355242> , 上網日期:2021/1/8.
- Krishna Dharavath, G. Amarnath, Fazal A. Talukdar and Rabul H. Laskar. (2014) *Impact of image preprocessing on face recognition: A comparative analysis*, IEEE Conference on International Conference on Communication and Signal Processing , pp. 630-635
- G. B. Huang, M. Ramesh, T. Berg and E. LearnedMiller, (2007). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*, Technical Report 07-49, University of Massachusetts, Amherst.
- Papageorgiou, Oren and Poggio. (1998). *A general framework for object detection* ,International Conference on Computer Vision.
- Viola and Jones. (2001). *Rapid object detection using a boosted cascade of simple features*, Computer Vision and Pattern Recognition.